



Adjoint and low-rank covariance representation

M. K. Tippett, S. E. Cohn

► To cite this version:

M. K. Tippett, S. E. Cohn. Adjoint and low-rank covariance representation. *Nonlinear Processes in Geophysics*, 2001, 8 (6), pp.331-340. hal-00302042

HAL Id: hal-00302042

<https://hal.science/hal-00302042>

Submitted on 1 Jan 2001

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Adjoint and low-rank covariance representation

M. K. Tippett¹ and S. E. Cohn²

¹International Research Institute for climate prediction, LDEO, Palisades, NY, USA

²Data Assimilation Office, NASA/GSFC, Greenbelt, MD, USA

Received: 10 October 2000 – Revised: 24 August 2001 – Accepted: 4 September 2001

Abstract. Quantitative measures of the uncertainty of Earth system estimates can be as important as the estimates themselves. Direct calculation of second moments of estimation errors, as described by the covariance matrix, is impractical when the number of degrees of freedom of the system state is large and the sources of uncertainty are not completely known. Theoretical analysis of covariance equations can help guide the formulation of low-rank covariance approximations, such as those used in ensemble and reduced-state approaches for prediction and data assimilation. We use the singular value decomposition and recently developed *positive map* techniques to analyze a family of covariance equations that includes stochastically forced linear systems. We obtain covariance estimates given imperfect knowledge of the sources of uncertainty and we obtain necessary conditions for low-rank approximations to be appropriate. The results are illustrated in a stochastically forced system with time-invariant linear dynamics.

1 Introduction

Estimates of the state of the Earth system contain uncertainty, a consequence of imperfect observations and models. Quantitative measures of this uncertainty are useful in a variety of ways. Users of weather and climate predictions base decisions on both forecast and forecast uncertainty (Changnon et al., 1999). Data assimilation systems combine information from observations and models in a manner that depends upon their presumed uncertainty (Cohn, 1997). Ensemble prediction systems use estimates of analysis error to generate ensembles of initial conditions (Barkmeijer et al., 1998).

Uncertainty can be modeled as a random variable. With some assumptions, equations can often be obtained for the evolution of the mean and covariance of this random variable (Ehrendorfer and Tribbia, 1997). For instance, the co-

variance of stochastically forced linear systems is described by a Lyapunov equation (Farrell and Ioannou, 1996). However, for dynamical systems with a large number of degrees of freedom, direct solution of covariance equations is impractical due to computational cost and incomplete knowledge of the sources of uncertainty. Therefore, in many situations, the covariance must be modeled.

One approach to covariance modeling is to specify analytically parameters, such as the variances and correlation lengths (Dee and da Silva, 1998; Rabier et al., 1998). Complex features, such as flow dependence, may also be modeled through appropriate parameterization (Riishøjgaard, 1998). A second approach is to assume that the uncertainty is well described with a few structures and hence, to approximate the covariance matrix by a low-rank matrix. Low-dimensional covariance representations are often directly connected to the dynamics, as in reduced-state Kalman filter data assimilation and ensemble prediction (Cane et al., 1996; Cohn and Todling, 1996; Houtekamer and Mitchell, 1998; Molteni et al., 1996). Correlation modeling and low-rank approaches are combined in “hybrid” methods (Hamill and Snyder, 2000).

Direct comparison of covariance approximations with the exact covariance is only possible in idealized models where the number of degrees of freedom is small and the sources of uncertainty are specified (Cohn and Todling, 1996; Kleeman and Moore, 1997; Whitaker and Sardeshmukh, 1998). Otherwise, analytical methods must be relied upon to provide guidance. Theoretical analysis of covariance equations can provide insights with potential value for covariance modeling. An example is Tippett et al. (2000b), where the steady-state analysis error covariance of a time-invariant data assimilation system is analyzed.

The purpose of the present paper is to analyze a more general problem, namely, the family of covariance equations whose solution is a linear transformation of a single forcing covariance. This family of problems includes systems with linear dynamics and stationary additive stochastic forcing. We use two analytical methods. The first uses the fa-

milar eigenvalue and singular value decompositions (SVD) and provides results for normal operators. These results are extended to the general nonnormal case, using a *positive map* technique that has recently been applied to Lyapunov equations (Bhatia, 1997; Tippett et al., 2000a). We demonstrate that many results derived for specific covariance equations are actually simple general consequences of the structure of positive maps. In particular, we show that some known properties of Lyapunov equations, such as covariance bounds, stochastic optimals and necessary conditions for low-rank approximation, are, in fact, properties of the general problem. Additionally, we obtain a new low-rank approximation estimate.

The paper is organized as follows. Section 2 introduces the general linear covariance equation. This equation is analyzed in Sect. 3, using the SVD and positive map techniques. Section 4 illustrates the results with an example, using the dynamics of a generalized nonnormal advection equation. Conclusions are given in Sect. 5.

2 The linear covariance equation

We suppose that the system uncertainty ϵ , a real vector of length n , is a mean-zero random variable. This ϵ might be a forecast error, an analysis error, or a climate anomaly, for instance. From physical and dimensional considerations, a symmetric positive definite matrix \mathbf{M} is chosen so that the scalar $\epsilon^T \mathbf{M} \epsilon$ is a meaningful measure of the magnitude of the uncertainty ϵ ; $(\cdot)^T$ denotes the transpose. The corresponding inner product for (generally complex) vectors is defined to be $(\epsilon_1, \epsilon_2) \equiv \epsilon_1^\dagger \mathbf{M} \epsilon_2$; $(\cdot)^\dagger$ denotes the conjugate transpose. We take $\mathbf{M} = \mathbf{I}$ without loss of generality since if $\mathbf{M} \neq \mathbf{I}$, then a new random variable $\hat{\epsilon} = \mathbf{M}^{1/2} \epsilon$ can be introduced with $(\epsilon_1, \epsilon_2) = \hat{\epsilon}_1^\dagger \hat{\epsilon}_2$. System uncertainty statistics are contained in the $n \times n$ covariance matrix \mathbf{P} , defined by

$$\mathbf{P} \equiv \langle \epsilon \epsilon^T \rangle, \quad (1)$$

where $\langle \cdot \rangle$ denotes expectation. The total variance of ϵ is given by $\text{tr } \mathbf{P}$, where tr denotes trace.

The eigenvectors or EOFs of \mathbf{P} order state-space directions according to the amount of variance they explain and can be used to construct low-rank approximations of \mathbf{P} . Rank- r approximations of \mathbf{P} , $r \ll n$, can be stored and used in calculations even when the size of n makes calculations with \mathbf{P} itself impractical. A particular rank- r approximation $\mathbf{P}^{(r)}$ is the projection of \mathbf{P} onto its leading r eigenvectors, given by

$$\mathbf{P}^{(r)} \equiv \sum_{i=1}^r \lambda_i(\mathbf{P}) \mathbf{w}_i \mathbf{w}_i^T, \quad (2)$$

where \mathbf{w}_i is the normalized eigenvector of \mathbf{P} corresponding to the eigenvalue $\lambda_i(\mathbf{P})$ with ordering $\lambda_1(\mathbf{P}) \geq \lambda_2(\mathbf{P}) \geq \dots \geq \lambda_n(\mathbf{P}) \geq 0$. This projection $\mathbf{P}^{(r)}$ is, in fact, the best rank- r approximation of \mathbf{P} in the sense that it minimizes the error $\|\mathbf{P} - \mathbf{P}^{(r)}\|$ for any unitarily invariant matrix norm $\|\cdot\|$

(Horn and Johnson, 1985). The eigenvalues of \mathbf{P} completely determine the approximation error, which is

$$\|\mathbf{P} - \mathbf{P}^{(r)}\| = \|\text{diag}(0, \dots, 0, \lambda_{r+1}(\mathbf{P}), \dots, \lambda_n(\mathbf{P}))\|. \quad (3)$$

For instance, in the spectral norm $\|\mathbf{X}\|_\infty \equiv \sqrt{\lambda_1(\mathbf{X}^T \mathbf{X})}$, the approximation error is

$$\begin{aligned} \|\mathbf{P} - \mathbf{P}^{(r)}\|_\infty &= \lambda_1(\text{diag}(0, \dots, 0, \lambda_{r+1}(\mathbf{P}), \dots, \lambda_n(\mathbf{P}))) \\ &= \lambda_{r+1}(\mathbf{P}). \end{aligned} \quad (4)$$

In this norm, the error is small compared with $\|\mathbf{P}\|_\infty$ when the first eigenvalue of \mathbf{P} is large compared to the $(r+1)$ st eigenvalue of \mathbf{P} . In fact, a necessary condition for \mathbf{P} to have a good low-rank approximation in *any* unitarily invariant norm is that $\lambda_1(\mathbf{P}) \gg \lambda_n(\mathbf{P})$, i.e. that \mathbf{P} has a large condition number $\kappa(\mathbf{P}) \equiv \|\mathbf{P}\|_\infty \|\mathbf{P}^{-1}\|_\infty$ (Golub and Van Loan, 1996); the condition number of a symmetric positive definite matrix \mathbf{P} is $\lambda_1(\mathbf{P})/\lambda_n(\mathbf{P})$. Such a matrix is said to be ill-conditioned.

Equations for the covariance \mathbf{P} can be derived from evolution equations for the system uncertainty ϵ . For instance, suppose the evolution of ϵ is given by

$$\begin{aligned} \frac{d\epsilon}{dt} &= \mathbf{F}\epsilon + \xi, \quad \epsilon(t=0) = 0, \\ \langle \xi(t_1) \xi(t_2)^T \rangle &= \delta(t_1 - t_2) \mathbf{P}_0, \end{aligned} \quad (5)$$

where \mathbf{F} is a constant $n \times n$ matrix. Then we obtain a Lyapunov equation for \mathbf{P} :

$$\frac{d\mathbf{P}}{dt} = \mathbf{F}\mathbf{P} + \mathbf{P}\mathbf{F}^T + \mathbf{P}_0, \quad \mathbf{P}(t=0) = 0. \quad (6)$$

Two basic properties of (6) are that its solution \mathbf{P} is a covariance matrix for any forcing covariance \mathbf{P}_0 , and that \mathbf{P} is a linear function of \mathbf{P}_0 . We will see that much of the structure of this solution is a consequence of these two simple properties.

To demonstrate this assertion as the basis for this paper, we consider a generalization of (6), namely, the linear covariance equation

$$\mathbf{P} = \mathcal{L}\mathbf{P}_0, \quad \mathbf{P}_0 = \mathbf{P}_0^T \geq 0, \quad (7)$$

where \mathcal{L} is an operator acting on matrices and \mathbf{P}_0 is a forcing covariance matrix. We assume only that (i) \mathcal{L} is such that \mathbf{P} is a covariance matrix for any forcing covariance \mathbf{P}_0 and that (ii) \mathbf{P} is a linear function of \mathbf{P}_0 , i.e. that \mathcal{L} is a linear operator. In the case of (6), these two properties can be established by writing its solution explicitly in the form of (7):

$$\mathbf{P}(t) = \mathcal{L}\mathbf{P}_0 \equiv \int_0^t e^{\tau \mathbf{F}} \mathbf{P}_0 e^{\tau \mathbf{F}^T} d\tau; \quad (8)$$

here \mathcal{L} depends on t . Similarly, if the evolution of the uncertainty is given by a discrete-time model

$$\begin{aligned} \epsilon_{k+1} &= \mathbf{A}\epsilon_k + \xi_{k+1}, \quad \langle \epsilon_0 \epsilon_0^T \rangle = \mathbf{P}_0, \\ \langle \xi_i \xi_j^T \rangle &= \delta_{ij} \mathbf{P}_0, \end{aligned} \quad (9)$$

then the covariance evolution is

$$\mathbf{P}_{k+1} = \mathbf{A}\mathbf{P}_k\mathbf{A}^T + \mathbf{P}_0. \quad (10)$$

The solution of (10) can be written in the form of (7) as

$$\mathbf{P}_{k+1} = \mathcal{L}\mathbf{P}_0 \equiv \sum_{i=0}^{k+1} \mathbf{A}^i \mathbf{P}_0 (\mathbf{A}^T)^i; \quad (11)$$

here \mathcal{L} depends on k . For stable¹ dynamics, the limits $t \rightarrow \infty$ and $k \rightarrow \infty$ of (8) and (11), respectively, exist and give the solutions of the algebraic Lyapunov equations:

$$\mathbf{F}\mathbf{P} + \mathbf{P}\mathbf{F}^T + \mathbf{P}_0 = 0 \quad (12)$$

and

$$\mathbf{P} = \mathbf{A}\mathbf{P}\mathbf{A}^T + \mathbf{P}_0. \quad (13)$$

As (8) and (11) show, in these cases \mathcal{L} is a highly nonlinear function of the dynamics and for this reason, its analysis is challenging.

We examine next how the two defining properties of \mathcal{L} in the linear covariance equation (7) determine the structure of the solution \mathbf{P} . Since the forcing covariance \mathbf{P}_0 is often poorly known, results that depend primarily on \mathcal{L} and require limited knowledge of \mathbf{P}_0 are desirable.

3 Analysis of linear covariance matrix equations

3.1 Eigenvectors and singular vectors

We present two methods of analyzing the solution of (7). In both, the fundamental object of our attention will be the linear operator \mathcal{L} that maps the covariance forcing \mathbf{P}_0 to the covariance response \mathbf{P} . In the first method, the covariance matrices \mathbf{P}_0 and \mathbf{P} are viewed as vectors of length n^2 and \mathcal{L} is viewed as an $n^2 \times n^2$ matrix (Byers and Nash, 1987; Ghavimi and Laub, 1995). Familiar linear algebra techniques, such as the SVD and the eigenvalue decomposition, can then be applied. For instance, the SVD of \mathcal{L} can be used to identify the forcing \mathbf{P}_0 that produces the maximum response \mathbf{P} , and can also be used to identify approximations of \mathcal{L} . The SVD of \mathcal{L} and the adjoint \mathcal{L}^* of \mathcal{L} depend on the choice of a *matrix inner product*.

A natural inner product for two matrices \mathbf{X} and \mathbf{Y} can be defined by

$$(\mathbf{X}, \mathbf{Y}) \equiv \text{tr } \mathbf{M}^{1/2} \mathbf{X}^\dagger \mathbf{M} \mathbf{Y} \mathbf{M}^{1/2}, \quad (14)$$

where \mathbf{M} is a symmetric positive definite matrix. The matrix inner product in (14) is compatible with the vector inner product $(\epsilon_1, \epsilon_2) = \epsilon_1^\dagger \mathbf{M} \epsilon_2$, defined previously in the sense that the orthogonality of two rank-1 Hermitian matrices $\epsilon_1 \epsilon_1^\dagger$ and $\epsilon_2 \epsilon_2^\dagger$ is equivalent to the orthogonality of the vectors ϵ_1

¹The continuous-time dynamics \mathbf{F} is stable if and only if the eigenvalues of \mathbf{F} all have a real part that is less than zero. The discrete-time dynamics \mathbf{A} is stable if and only if the eigenvalues of \mathbf{A} lie inside the unit circle.

and ϵ_2 . Again, we take $\mathbf{M} = \mathbf{I}$ without a loss of generality. We mention that for $\mathbf{M} = \mathbf{I}$, the matrix inner product (14) is just the Euclidean inner product on vectors of length n^2 , since then

$$(\mathbf{X}, \mathbf{Y}) = \text{tr } \mathbf{X}^\dagger \mathbf{Y} = \sum_{i=1}^n \sum_{j=1}^n \bar{\mathbf{X}}_{ij} \mathbf{Y}_{ij} \quad (15)$$

and that $(\mathbf{X}, \mathbf{X}) = \|\mathbf{X}\|_2^2$, where $\|\cdot\|_2$ is the Frobenius matrix norm.

By definition, the adjoint \mathcal{L}^* of \mathcal{L} satisfies

$$(\mathbf{X}, \mathcal{L}\mathbf{Y}) = (\mathcal{L}^*\mathbf{X}, \mathbf{Y}), \quad (16)$$

for any two $n \times n$ matrices \mathbf{X} and \mathbf{Y} . It can be shown, for instance, that the adjoints of the operators \mathcal{L} in (8) and (11) with respect to the inner product (15) are, respectively,

$$\mathcal{L}^*\mathbf{P}_0 \equiv \int_0^t e^{\tau\mathbf{F}^T} \mathbf{P}_0 e^{\tau\mathbf{F}} d\tau \quad (17)$$

and

$$\mathcal{L}^*\mathbf{P}_0 \equiv \sum_{i=0}^{k+1} (\mathbf{A}^T)^i \mathbf{P}_0 \mathbf{A}^i. \quad (18)$$

These two covariance operators \mathcal{L} are *normal*, i.e. they have a complete set of orthogonal eigenvectors, if and only if the dynamics (\mathbf{F} and \mathbf{A} , respectively) are normal.

Suppose we denote by $\sigma_i(\mathcal{L})$, \mathbf{U}_i and \mathbf{V}_i , respectively, the i -th singular value, left singular vector and right singular vector of the general operator \mathcal{L} , $1 \leq i \leq n^2$. The $n \times n$ matrices \mathbf{U}_i and \mathbf{V}_i satisfy $\mathcal{L}\mathbf{V}_i = \sigma_i(\mathcal{L})\mathbf{U}_i$ and $(\mathbf{U}_i, \mathbf{U}_j) = (\mathbf{V}_i, \mathbf{V}_j) = \delta_{ij}$, where (\cdot, \cdot) is the matrix inner product (15); singular values are ordered so that $\sigma_1(\mathcal{L}) \geq \sigma_2(\mathcal{L}) \geq \dots \geq \sigma_{n^2}(\mathcal{L}) \geq 0$. Then the decomposition of the solution \mathbf{P} of the covariance equation (7) in the left singular vectors of \mathcal{L} is

$$\mathbf{P} = \mathcal{L}\mathbf{P}_0 = \sum_{i=1}^{n^2} \sigma_i(\mathcal{L}) (\mathbf{V}_i, \mathbf{P}_0) \mathbf{U}_i. \quad (19)$$

The first singular value of \mathcal{L} is seen to be the maximum amplification of the forcing \mathbf{P}_0 in the Frobenius norm, i.e.

$$\max_{\mathbf{P}_0} \frac{\|\mathcal{L}\mathbf{P}_0\|_2}{\|\mathbf{P}_0\|_2} = \sigma_1(\mathcal{L}) \quad (20)$$

and is achieved by choosing $\mathbf{P}_0 = \mathbf{V}_1$, in which case $\mathbf{P} = \sigma_1(\mathcal{L})\mathbf{U}_1$.

Approximate solutions of the covariance equation (7) can be obtained by approximating \mathcal{L} . Using the approximation $\mathcal{L}^{(r)}$ obtained by truncating the series in (19) gives

$$\tilde{\mathbf{P}} \equiv \mathcal{L}^{(r)}\mathbf{P}_0 \equiv \sum_{i=1}^r \sigma_i(\mathcal{L}) (\mathbf{V}_i, \mathbf{P}_0) \mathbf{U}_i \quad (21)$$

with error

$$\begin{aligned} \|\mathbf{P} - \tilde{\mathbf{P}}\|_2^2 &= \left\| \left(\mathcal{L} - \mathcal{L}^{(r)} \right) \mathbf{P}_0 \right\|_2^2 \\ &= \sum_{i=r+1}^{n^2} \sigma_i^2(\mathcal{L}) |(\mathbf{V}_i, \mathbf{P}_0)|^2. \end{aligned} \quad (22)$$

The error of this approximation depends on the size of the singular values of \mathcal{L} and on the projection of \mathbf{P}_0 onto the right singular vectors \mathbf{V}_i . The approximation error is small relative to $\|\mathbf{P}\|_2$ if the leading r singular values of \mathcal{L} are large compared to $\sigma_{r+1}(\mathcal{L})$, and if the projection of \mathbf{P}_0 onto the leading r right singular vectors \mathbf{V}_i , $1 \leq i \leq r$, is not small. If the leading r left singular vectors \mathbf{U}_i of \mathcal{L} happen to be rank-1 matrices, then $\tilde{\mathbf{P}}$ is an approximation of \mathbf{P} whose rank is at most r , although this $\tilde{\mathbf{P}}$ may not be the best rank- r approximation of $\mathbf{P}^{(r)}$ defined in (2). Still, in case the leading r left singular vectors \mathbf{U}_i of \mathcal{L} happen to be rank-1 matrices, we have constructed a low-rank approximation $\tilde{\mathbf{P}}$ of \mathbf{P} whose error depends on the singular values of \mathcal{L} .

To illustrate a situation where the rank of $\tilde{\mathbf{P}}$ in (21) is at most r , consider the solution \mathbf{P} of the discrete algebraic Lyapunov equation (13) for normal dynamics \mathbf{A} ,

$$\mathbf{P} = \mathcal{L}_{\mathbf{A}} \mathbf{P}_0 \equiv \sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{P}_0 (\mathbf{A}^T)^k; \quad (23)$$

the notation $\mathcal{L}_{\mathbf{A}}$ emphasizes the dependence on the dynamics \mathbf{A} . The eigenvalues and eigenvectors of the Lyapunov operator $\mathcal{L}_{\mathbf{A}}$ in (23) are simply related to those of the dynamics \mathbf{A} (Lancaster, 1970). Specifically, $D_{ij} \equiv (1 - \lambda_i(\mathbf{A})\overline{\lambda_j(\mathbf{A})})^{-1}$ is an eigenvalue of $\mathcal{L}_{\mathbf{A}}$ whose corresponding eigenvector is the rank-1 matrix $\mathbf{z}_i \mathbf{z}_j^\dagger$, where \mathbf{z}_i is the eigenvector of \mathbf{A} corresponding to the eigenvalue $\lambda_i(\mathbf{A})$, i.e.

$$\mathcal{L}_{\mathbf{A}} \mathbf{z}_i \mathbf{z}_j^\dagger = \left(1 - \lambda_i(\mathbf{A})\overline{\lambda_j(\mathbf{A})}\right)^{-1} \mathbf{z}_i \mathbf{z}_j^\dagger. \quad (24)$$

When \mathbf{A} is normal, so is $\mathcal{L}_{\mathbf{A}}$, and the eigenvectors of $\mathcal{L}_{\mathbf{A}}$ are orthogonal with respect to the inner product (15). In this case, $\mathcal{L}_{\mathbf{A}}^* \mathbf{z}_i \mathbf{z}_j^\dagger = \overline{D_{ij}} \mathbf{z}_i \mathbf{z}_j^\dagger$ and it follows that $\mathcal{L}_{\mathbf{A}} \mathcal{L}_{\mathbf{A}}^* \mathbf{z}_i \mathbf{z}_j^\dagger = \mathcal{L}_{\mathbf{A}}^* \mathcal{L}_{\mathbf{A}} \mathbf{z}_i \mathbf{z}_j^\dagger = |D_{ij}|^2 \mathbf{z}_i \mathbf{z}_j^\dagger$. Therefore, $|D_{ij}|$ is a singular value of the normal operator $\mathcal{L}_{\mathbf{A}}$. If D_{ij} is real, then the rank-1 matrix $\mathbf{z}_i \mathbf{z}_j^\dagger$ is the corresponding singular vector. If D_{ij} is complex, then $|D_{ij}|$ is a repeated singular value and the corresponding singular vectors are the rank-2 matrices $\Re \mathbf{z}_i \mathbf{z}_j^\dagger$ and $\Im \mathbf{z}_i \mathbf{z}_j^\dagger$, whose sum has rank-2. Therefore, the matrix $\tilde{\mathbf{P}}$ in (21) has rank which is at most r when the truncation is chosen such that $\sigma_{r+1}(\mathcal{L}_{\mathbf{A}}) \neq \sigma_r(\mathcal{L}_{\mathbf{A}})$.

When \mathbf{A} is normal, SVD analysis of $\mathcal{L}_{\mathbf{A}}$ can be replaced by eigenanalysis. The expansion of \mathbf{P} in the eigenvectors of $\mathcal{L}_{\mathbf{A}}$ is

$$\begin{aligned} \mathbf{P} &= \sum_{i=1}^n \sum_{j=1}^n D_{ij} \operatorname{tr}(\mathbf{z}_i \mathbf{z}_j^\dagger \mathbf{P}_0) \mathbf{z}_i \mathbf{z}_j^\dagger \\ &= \sum_{i=1}^n \sum_{j=1}^n D_{ij} (\mathbf{z}_j^\dagger \mathbf{P}_0 \mathbf{z}_i) \mathbf{z}_i \mathbf{z}_j^\dagger. \end{aligned} \quad (25)$$

The maximum possible amplification in the Frobenius norm, as well as in any other unitarily invariant norm, is given by the first eigenvalue $\lambda_1(\mathcal{L}_{\mathbf{A}}) = (1 - |\lambda_1(\mathbf{A})|^2)^{-1}$ and is achieved by the forcing $\mathbf{P}_0 = \mathbf{z}_1 \mathbf{z}_1^\dagger$, for which $\mathbf{P} = \mathcal{L}_{\mathbf{A}} \mathbf{P}_0 = (1 - |\lambda_1(\mathbf{A})|^2)^{-1} \mathbf{z}_1 \mathbf{z}_1^\dagger$. A rank- r approximation of \mathbf{P} is ob-

tained by truncating the series expansion (25):

$$\tilde{\mathbf{P}}^{(r)} = \sum_{i=1}^r \sum_{j=1}^r D_{ij} (\mathbf{z}_j^\dagger \mathbf{P}_0 \mathbf{z}_i) \mathbf{z}_i \mathbf{z}_j^\dagger. \quad (26)$$

The error of this approximation depends on (i) the values of D_{ij} , which, in turn, are determined by the stability properties of the dynamics \mathbf{A} , and on (ii) how the forcing covariance \mathbf{P}_0 projects onto the eigenvectors of \mathbf{A} . When $\mathbf{P}_0 = \mathbf{I}$, the expansion (25) becomes simply

$$\mathbf{P} = \sum_{i=1}^n D_{ii} \mathbf{z}_i \mathbf{z}_i^\dagger, \quad (27)$$

and

$$\tilde{\mathbf{P}}^{(r)} = \sum_{i=1}^r D_{ii} \mathbf{z}_i \mathbf{z}_i^\dagger \quad (28)$$

is, in fact, the best rank- r approximation $\mathbf{P}^{(r)}$ of \mathbf{P} . For $\mathbf{P}_0 = \mathbf{I}$, the error $\|\mathbf{P} - \tilde{\mathbf{P}}^{(r)}\|_2$ is small relative to $\|\mathbf{P}\|_2$ when some, but not all of the eigenmodes of \mathbf{A} are nearly unstable, i.e. when some, but not all of the D_{ii} are large. In this case, $\lambda_1(\mathcal{L}_{\mathbf{A}}) \gg \lambda_{n^2}(\mathcal{L}_{\mathbf{A}})$ and $\mathcal{L}_{\mathbf{A}}$ is ill-conditioned. When $\mathbf{P}_0 \neq \mathbf{I}$, a rank- r approximation of \mathbf{P} may be better than (26) if the spectrum of $\mathcal{L}_{\mathbf{A}}$ is relatively flat and hence, $\mathcal{L}_{\mathbf{A}}$ is not ill-conditioned. For instance, when $\mathbf{A} = c\mathbf{I}$, $0 \leq c < 1$, the spectrum of $\mathcal{L}_{\mathbf{A}}$ is flat and $\lambda_1(\mathcal{L}_{\mathbf{A}}) = \lambda_{n^2}(\mathcal{L}_{\mathbf{A}}) = (1 - c^2)^{-1}$. In this case, the solution \mathbf{P} for any \mathbf{P}_0 is given by

$$\mathbf{P} = \frac{1}{1 - c^2} \mathbf{P}_0 \quad (29)$$

and has the best rank- r approximation $\mathbf{P}^{(r)} = (1 - c^2)^{-1} \mathbf{P}_0^{(r)}$.

The examples (27) and (29) illustrate two factors that may lead to the solution \mathbf{P} of the discrete algebraic Lyapunov equation (13) for normal dynamics having a good low-rank approximation: $\mathcal{L}_{\mathbf{A}}$ being ill-conditioned and \mathbf{P}_0 having a good low-rank approximation. Recall that a necessary condition for a positive semi-definite matrix \mathbf{P} to have a good low-rank approximation is that it be ill-conditioned, $\lambda_1(\mathbf{P}) \gg \lambda_n(\mathbf{P})$. In fact, the solution \mathbf{P} of (13) can be ill-conditioned only when either $\mathcal{L}_{\mathbf{A}}$ or \mathbf{P}_0 is ill-conditioned. To observe this for normal dynamics, one writes (25) as

$$\mathbf{P} = \mathbf{Z} (\mathbf{D} \circ (\mathbf{Z}^\dagger \mathbf{P}_0 \mathbf{Z})) \mathbf{Z}^\dagger, \quad (30)$$

where \circ denotes the Hadamard product² and $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_n]$. Then one can obtain³

$$\frac{\lambda_1(\mathbf{P})}{\lambda_n(\mathbf{P})} \leq \frac{\max_i D_{ii} \lambda_1(\mathbf{P}_0)}{\min_i D_{ii} \lambda_n(\mathbf{P}_0)} \leq \frac{\lambda_1(\mathcal{L}_{\mathbf{A}}) \lambda_1(\mathbf{P}_0)}{\lambda_{n^2}(\mathcal{L}_{\mathbf{A}}) \lambda_n(\mathbf{P}_0)} \quad (31)$$

²The Hadamard product of two matrices \mathbf{X} and \mathbf{Y} with entries X_{ij} and Y_{ij} , respectively, is the matrix whose entries are $X_{ij}Y_{ij}$.

³From $\lambda_n(\mathbf{P}_0)(\mathbf{D} \circ \mathbf{I}) \leq \mathbf{D} \circ \mathbf{P}_0 \leq \lambda_1(\mathbf{P}_0)(\mathbf{D} \circ \mathbf{I})$.

or equivalently, $\kappa(\mathbf{P}) \leq \kappa(\mathcal{L}_A)\kappa(\mathbf{P}_0)$, showing that \mathbf{P} can be ill-conditioned only if either \mathcal{L}_A or \mathbf{P}_0 is also ill-conditioned. These two mechanisms can interfere with each other and examples can be constructed where both \mathcal{L}_A and \mathbf{P}_0 are ill-conditioned, but the spectrum of \mathbf{P} is flat.

There is no correspondingly simple analysis of the solution of (13) for nonnormal dynamics \mathbf{A} . In general, the singular vectors of \mathcal{L}_A are rank- n matrices and the truncation in (21) does not give a low-rank approximation of \mathbf{P} . Additionally, there is the practical difficulty of the calculation of the SVD of \mathcal{L}_A . However, in the next subsection, we show that the properties of the normal Lyapunov operator that lead to relations such as (31) are also properties of both the nonnormal Lyapunov operator and of general \mathcal{L} . This allows us to show for the Lyapunov operator, a relation between the stability of \mathbf{A} and the conditioning of \mathbf{P} , and for general \mathcal{L} , to show the connection between the conditioning of \mathcal{L} and that of \mathbf{P} .

3.2 Positive maps and operator norms

The linear operator \mathcal{L} in (7) maps covariance matrices to covariance matrices and is thus, a *positive map*. We will use the properties of positive maps to extend the results of the previous section and to demonstrate that many properties of solutions of Lyapunov equations are also properties of solutions of (7) (Bhatia, 1997; Tippett et al., 2000a). We use *operator norms* to obtain a new upper bound for the fraction of variance explained by the first eigenmode of \mathbf{P} .

The operator norm of \mathcal{L} is defined by

$$\|\mathcal{L}\|_p \equiv \max_{\mathbf{X}} \frac{\|\mathcal{L}\mathbf{X}\|_p}{\|\mathbf{X}\|_p}; \quad (32)$$

the Schatten p -norm of an $n \times n$ matrix \mathbf{X} is defined to be

$$\|\mathbf{X}\|_p \equiv \left(\sum_{i=1}^n \sigma_i^p(\mathbf{X}) \right)^{1/p} \quad 1 \leq p \leq \infty, \quad (33)$$

where $\sigma_i(\mathbf{X})$ is the i -th singular value of the matrix \mathbf{X} . For covariance matrices \mathbf{P} , $\|\mathbf{P}\|_1 = \text{tr } \mathbf{P}$, $\|\mathbf{P}\|_2$ is the Frobenius norm, and $\|\mathbf{P}\|_\infty = \lambda_1(\mathbf{P})$. For normal dynamics \mathbf{A} , the norm of the Lyapunov operator is $\|\mathcal{L}_A\|_p = \lambda_1(\mathcal{L}_A) = (1 - |\lambda_1(\mathbf{A})|^2)^{-1}$.

A basic fact about positive maps is that they obtain their $p = \infty$ Schatten norm on the identity matrix \mathbf{I} (Bhatia, 1997), i.e.

$$\|\mathcal{L}\|_\infty \equiv \max_{\mathbf{P}_0} \frac{\lambda_1(\mathcal{L}\mathbf{P}_0)}{\lambda_1(\mathbf{P}_0)} = \lambda_1(\mathbf{B}), \quad (34)$$

where the *bound matrix* \mathbf{B} , defined by $\mathbf{B} \equiv \mathcal{L}\mathbf{I}$, is the covariance response (7) to the identity. This means that rather than having to carry out the maximization in definition (32), $\|\mathcal{L}\|_\infty$ can always be calculated by computing the largest eigenvalue of the bound matrix. The bound matrix \mathbf{B} can also be used to estimate the variance explained by a direction \mathbf{x} in state space by using the relation

$$\lambda_n(\mathbf{P}_0) \mathbf{x}^T \mathbf{B} \mathbf{x} \leq \mathbf{x}^T \mathbf{P} \mathbf{x} \leq \lambda_1(\mathbf{P}_0) \mathbf{x}^T \mathbf{B} \mathbf{x}, \quad (35)$$

derived in Tippett and Marchesin (1999) for the discrete algebraic Lyapunov equation. In this sense, the eigenvectors of the bound matrix order state space directions according to the maximum possible response. Sometimes when \mathbf{P}_0 is poorly known, \mathbf{P} is calculated assuming $\mathbf{P}_0 = \mathbf{I}$ (Whitaker and Sardeshmukh, 1998). The bounds in (35) show that the results of such calculations can be used to bound the response to a general forcing covariance \mathbf{P}_0 .

A convenient characterization of the $p = 1$ Schatten norm of \mathcal{L} is useful, since this norm measures the total variance, $\|\mathbf{P}\|_1 = \text{tr } \mathbf{P}$. Since the $p = 1$ and $p = \infty$ Schatten norms are dual to each other, a standard functional analysis result is that $\|\mathcal{L}\|_1 = \|\mathcal{L}^*\|_\infty$, where \mathcal{L}^* is the adjoint of \mathcal{L} with respect to the matrix norm defined in (15). Since \mathcal{L}^* is also a positive map, (34) can be used to compute the maximum total variance amplification:

$$\begin{aligned} \|\mathcal{L}\|_1 &\equiv \max_{\mathbf{P}_0} \frac{\text{tr } \mathcal{L}\mathbf{P}_0}{\text{tr } \mathbf{P}_0} = \|\mathcal{L}^*\|_\infty = \|\mathcal{L}^*\mathbf{I}\|_\infty \\ &= \lambda_1(\mathcal{L}^*\mathbf{I}). \end{aligned} \quad (36)$$

In other words, $\|\mathcal{L}\|_1$ can be calculated by computing the largest eigenvalue of $\mathcal{L}^*\mathbf{I}$. The forcing \mathbf{P}_0 that produces this maximum amplification can be found by considering the response to the rank-1 forcing $\mathbf{P}_0 = \mathbf{w}\mathbf{w}^T$, where \mathbf{w} is a column vector of length n . The amplification of this forcing is

$$\begin{aligned} \frac{\|\mathcal{L}\mathbf{w}\mathbf{w}^T\|_1}{\|\mathbf{w}\mathbf{w}^T\|_1} &= \frac{\text{tr } \mathcal{L}\mathbf{w}\mathbf{w}^T}{\text{tr } \mathbf{w}\mathbf{w}^T} = \frac{(\mathcal{L}^*\mathbf{I}, \mathbf{w}\mathbf{w}^T)}{\mathbf{w}^T \mathbf{w}} \\ &= \frac{\mathbf{w}^T (\mathcal{L}^*\mathbf{I}) \mathbf{w}}{\mathbf{w}^T \mathbf{w}}. \end{aligned} \quad (37)$$

The maximum value of the Rayleigh quotient on the right-hand side of (36) is $\lambda_1(\mathcal{L}^*\mathbf{I})$, obtained when \mathbf{w} is the leading eigenvector of $\mathcal{L}^*\mathbf{I}$. From (36), this is, in fact, the maximum variance amplification for any \mathbf{P}_0 . The eigenvectors of $\mathcal{L}^*\mathbf{I}$ are *stochastic optimals* in the sense that they order the state space directions according to the amount of variance excited by the forcing in that direction (Farrell and Ioannou, 1996; Kleeman and Moore, 1997). More generally, there is the relation $\text{tr } \mathbf{P} = \text{tr } (\mathbf{P}_0 \mathcal{L}^*\mathbf{I})$ (Bhatia, 1997). As an example, for the continuous-time Lyapunov equation with \mathcal{L} defined in (8), the stochastic optimals are the eigenvectors of

$$\mathcal{L}^*\mathbf{I} \equiv \int_0^t e^{\tau \mathbf{F}^T} e^{\tau \mathbf{F}} d\tau. \quad (38)$$

The positive map abstraction can also be used to formulate and prove properties of the solution \mathbf{P} of the general covariance equation (7) that are not obvious from analysis of a specific covariance equation. For instance, positive maps and operator norms are used in Tippett et al. (2000a) to formulate and prove the generalization of the bound (31):

$$\frac{\lambda_1(\mathbf{P})}{\lambda_n(\mathbf{P})} \leq \|\mathcal{L}\|_\infty \|\mathcal{L}^{-1}\|_\infty \frac{\lambda_1(\mathbf{P}_0)}{\lambda_n(\mathbf{P}_0)} \quad (39)$$

for invertible positive maps \mathcal{L} and $\mathbf{P}_0 = \mathbf{P}_0^T > 0$. This bound means that the solution \mathbf{P} of (7) can be ill-conditioned only if either \mathcal{L} or \mathbf{P}_0 is also ill-conditioned. For example,

if $\mathbf{P}_0 = \mathbf{I}$, then \mathbf{P} can be ill-conditioned and have a good low-rank approximation only if \mathcal{L} is ill-conditioned.

The general result (39) can be applied to the discrete algebraic Lyapunov equation (13) in order to demonstrate the connection between the stability of general nonnormal dynamics \mathbf{A} and the conditioning of the solution \mathbf{P} . This is done by relating the stability of \mathbf{A} to the operator norms of $\mathcal{L}_\mathbf{A}$ and its inverse. First, $\|\mathcal{L}_\mathbf{A}^{-1}\|_p$ can be estimated in terms of the singular values of \mathbf{A} by

$$\max_i |1 - \sigma_i^2(\mathbf{A})| \leq \|\mathcal{L}_\mathbf{A}^{-1}\|_p \leq 1 + \sigma_1^2(\mathbf{A}). \quad (40)$$

Therefore, $\|\mathcal{L}_\mathbf{A}^{-1}\|_\infty \gg 1$ when $\sigma_1(\mathbf{A}) \gg 1$ and the dynamics presents strong nonmodal growth. Second, the size of $\|\mathcal{L}_\mathbf{A}\|_p$ is bounded by

$$\frac{1}{2r(\mathbf{A}) + r^2(\mathbf{A})} \leq \|\mathcal{L}_\mathbf{A}\|_p \leq \frac{1}{r^2(\mathbf{A})}, \quad (41)$$

where the *radius of stability*⁴ $r(\mathbf{A})$ is the distance from \mathbf{A} to the closest unstable matrix (Mori, 1990; Tippett et al., 2000a); $\|\mathcal{L}_\mathbf{A}\|_p$ is large when \mathbf{A} is close to an unstable matrix. Together, estimates (40) and (41) show that $\|\mathcal{L}_\mathbf{A}\|_\infty \|\mathcal{L}_\mathbf{A}^{-1}\|_\infty$ is large when the dynamics \mathbf{A} is nearly unstable and has at least one singular value that is not near unity.

A new bound for the fraction of total variance explained by the first eigenmode of \mathbf{P} is

$$\frac{\lambda_1(\mathbf{P})/\text{tr } \mathbf{P}}{\lambda_1(\mathbf{P}_0)/\text{tr } \mathbf{P}_0} = \frac{\|\mathcal{L}\mathbf{P}_0\|_\infty \|\mathcal{L}^{-1}\mathbf{P}\|_1}{\|\mathbf{P}_0\|_\infty \|\mathbf{P}\|_1} \leq \|\mathcal{L}\|_\infty \|\mathcal{L}^{-1}\|_1. \quad (42)$$

Though not obvious from specific examples of positive maps \mathcal{L} , the result follows directly from the operator norm definition. A significant fraction of the total variance of \mathbf{P} can be in its first eigenmode only if either the same is true of \mathbf{P}_0 , or if the quantity $\|\mathcal{L}\|_\infty \|\mathcal{L}^{-1}\|_1$ is large. The estimates (40) and (41) show that $\|\mathcal{L}_\mathbf{A}\|_\infty \|\mathcal{L}_\mathbf{A}^{-1}\|_1$ is large if \mathbf{A} is nearly unstable and has at least one singular value that is not near unity.

Positive maps preserve ordering⁵. This property is useful in case \mathbf{P}_0 is not precisely known, but there are upper and lower bounds $\mathbf{P}_0^- \leq \mathbf{P}_0 \leq \mathbf{P}_0^+$. In this case, upper and lower bounds for the solution \mathbf{P} of (7) are

$$\mathcal{L}\mathbf{P}_0^- \leq \mathbf{P} \leq \mathcal{L}\mathbf{P}_0^+. \quad (43)$$

From this relationship follows bounds for the eigenvalues, diagonal and total variance of \mathbf{P} :

$$\lambda_i(\mathcal{L}\mathbf{P}_0^-) \leq \lambda_i(\mathbf{P}) \leq \lambda_i(\mathcal{L}\mathbf{P}_0^+) \quad (44)$$

$$\text{diag}(\mathcal{L}\mathbf{P}_0^-) \leq \text{diag}(\mathbf{P}) \leq \text{diag}(\mathcal{L}\mathbf{P}_0^+) \quad (45)$$

$$\text{tr } \mathcal{L}\mathbf{P}_0^- \leq \text{tr } \mathbf{P} \leq \text{tr } \mathcal{L}\mathbf{P}_0^+. \quad (46)$$

⁴For normal stable matrices $r(\mathbf{A}) = 1 - |\lambda_1(\mathbf{A})|$, the distance from its largest eigenvalue to the unit circle. For nonnormal dynamics, the radius of stability depends on the pseudospectrum of \mathbf{A} (Trefethen, 1997). Eigenvalues near the unit circle, large singular values and sensitive eigenvalues cause the radius of stability to be small.

⁵For two symmetric matrices \mathbf{X} and \mathbf{Y} , the ordering $\mathbf{X} \leq \mathbf{Y}$ means that $\mathbf{Y} - \mathbf{X}$ is positive semi-definite.

Taking $\mathbf{P}_0^- = \lambda_n(\mathbf{P}_0)\mathbf{I}$ and $\mathbf{P}_0^+ = \lambda_1(\mathbf{P}_0)\mathbf{I}$ in (43) gives bounds that depend on the bound matrix:

$$\lambda_n(\mathbf{P}_0)\mathbf{B} \leq \mathbf{P} \leq \lambda_1(\mathbf{P}_0)\mathbf{B}. \quad (47)$$

When the bounds in (47) are tight and \mathbf{P} has a well separated set of leading eigenvalues, the leading eigenvectors of \mathbf{P} and \mathbf{B} span approximately the same subspaces (Golub and Van Loan, 1996, Theorem 7.2.4).

The numerical cost of calculating operator norms can be comparable to calculating the full covariance, though the issue of poorly known sources of uncertainty is avoided. Lanczos methods can be used to calculate the leading eigenvalues and eigenvectors of $\mathcal{L}\mathbf{I}$ when $\mathcal{L}\mathbf{I}$, or an approximation thereof, is available as an operator. For instance, in the case of the discrete algebraic Lyapunov equation, the approximation

$$\mathcal{L}_\mathbf{A}\mathbf{I} \approx \mathbf{I} + \mathbf{A}\mathbf{A}^T + \mathbf{A}^2\mathbf{A}^{2T} + \dots + \mathbf{A}^k\mathbf{A}^{kT} \quad (48)$$

can be applied to a vector if \mathbf{A} and \mathbf{A}^T are available as operators.

4 Example

We now illustrate the results with a specific example: a discrete algebraic Lyapunov equation. The dynamics come from the generalized one-dimensional advection equation

$$\epsilon_t + a\epsilon_x + c'(x)\epsilon = 0 \quad 0 \leq x \leq 1 \quad (49)$$

with $a > 0$, and the initial and boundary conditions

$$\epsilon(x, t = 0) = \epsilon_0(x), \quad \epsilon(x = 0, t) = 0, \quad (50)$$

respectively; the function $c(x)$ is assumed to be monotone decreasing for convenience. Nonnormality is due to the undifferentiated term $c'(x)\epsilon$ and the boundary condition. Similar dynamics are used in Chang et al. (2001) to model tropical Atlantic variability. This model, with periodic boundary conditions, is studied in Tippett et al. (2000b) in the context of data assimilation. We discuss first properties of the deterministic dynamics.

We define \mathbf{A}_τ to be the operator that advances the solution τ time units for some fixed τ , i.e. $\epsilon(x, t + \tau) = \mathbf{A}_\tau \epsilon(x, t)$. The dynamics operator \mathbf{A}_τ is given by (see Appendix)

$$\mathbf{A}_\tau \epsilon(x, t) = \begin{cases} 0 & 0 \leq x \leq a\tau \\ s(x)\epsilon(x - a\tau, t) & a\tau \leq x \leq 1 \end{cases}, \quad (51)$$

where $s(x) \equiv \exp[(c(x - a\tau) - c(x))/a]$; $s(x) > 1$ since we have taken $c(x)$ to be decreasing, and $s(x) \approx \exp(-\tau c'(x))$ if $a\tau \ll 1$. Disturbances move from left to right with speed a , growing at rate $s(x)$. Spatially discrete, exact dynamics can be constructed on the regular grid $\{x_1, x_2, \dots, x_n\}$, $x_i = i/n$, for integer Courant number $C \equiv a\tau n$, by evaluating (51) at the grid points; the result is our $n \times n$ dynamics matrix \mathbf{A} (see Appendix). The zero boundary condition at the left boundary forces the solution

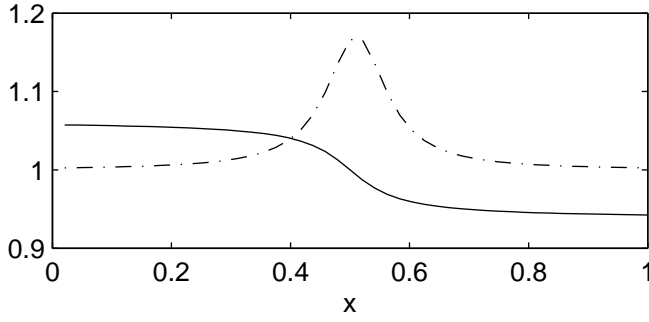


Fig. 1. $c(x)$ (solid line) and $s(x)$ (dash-dot line).

of (49) to be identically zero after the time a^{-1} required to cross the domain. Therefore, $\mathbf{A}^{1/\tau a} = 0$, and the matrix \mathbf{A} is nilpotent, with all of its eigenvalues identically zero; there is no modal growth.

Nonmodal transient growth is found from the singular values and singular vectors of \mathbf{A} . The singular values and left singular vectors of \mathbf{A} are the square roots of the eigenvalues and the eigenvectors of $\mathbf{A}\mathbf{A}^T$, which turns out to be a diagonal matrix (see Appendix):

$$(\mathbf{A}\mathbf{A}^T)_{ii} = \begin{cases} 0 & 1 \leq i \leq C \\ s^2(x_i) & C+1 \leq i \leq n. \end{cases} \quad (52)$$

Thus, the singular values of \mathbf{A} are zero and the values taken on by $s(x_i)$ at all but the first C grid points. The left singular vector associated with the singular value $s(x_i)$ is the i -th column of the identity matrix. The matrix $\mathbf{A}^T\mathbf{A}$ is also diagonal and the right singular vector associated with the singular value $s(x_i)$ is the $(i - C)$ -th column of the identity matrix.

We add mean-zero Gaussian-distributed noise ξ_k at each time-step:

$$\epsilon_{k+1} = \mathbf{A}\epsilon_k + \xi_k, \quad \langle \xi_i \xi_j^T \rangle = \delta_{ij} \mathbf{P}_0. \quad (53)$$

The steady-state covariance matrix \mathbf{P} of the system is given by $\mathbf{P} = \mathcal{L}_\mathbf{A} \mathbf{P}_0$. We have seen that many of the properties of $\mathcal{L}_\mathbf{A}$ can be obtained from the eigenvalues and eigenvectors of the matrices $\mathbf{B} \equiv \mathcal{L}_\mathbf{A} \mathbf{I}$ and $\mathbf{B}_T \equiv \mathcal{L}_\mathbf{A}^T \mathbf{I}$. These matrices are diagonal for the dynamics here. The largest response, as measured by the eigenvalues of \mathbf{B} , is $\|\mathcal{L}_\mathbf{A}\|_\infty = \lambda_1(\mathbf{B})$, according to (34). The largest sensitivity to forcing, as measured by the eigenvalues of \mathbf{B}_T , is $\|\mathcal{L}_\mathbf{A}\|_1 = \lambda_1(\mathbf{B}_T)$, according to (36). In fact, the maximum of $\text{diag } \mathbf{B}$ is within $a\tau$ of the right boundary and the maximum of $\text{diag } \mathbf{B}_T$ occurs within $a\tau$ of the left boundary, independently of $c(x)$ (see Appendix). In contrast, maximum growth as measured by singular vectors depends on $c(x)$ and is located at the maximum of $s(x)$. Since left singular vectors of \mathbf{A} are to the right of their corresponding right singular vectors, the leading left singular vectors of \mathbf{A} explain more of the total variance of \mathbf{B} than the leading right singular vectors of \mathbf{A} .

As a specific numerical example, we take $a = 1/12$, $\tau =$

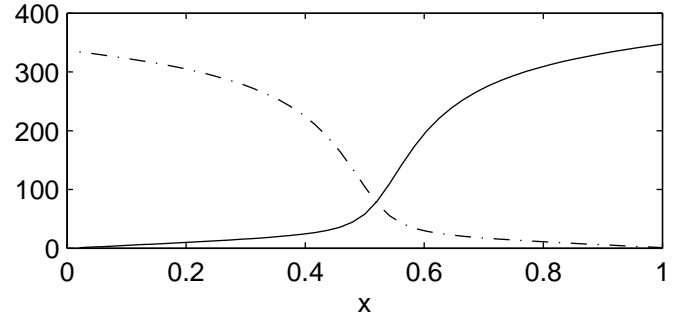


Fig. 2. The diagonal elements of \mathbf{B} (solid line) and \mathbf{B}_T (dash-dot line).

0.25, $n = 48$, and $c(x)$ given by

$$c(x) = 1 + \frac{1}{16} \frac{2}{\pi} \arctan(16(0.5 - x)). \quad (54)$$

In much of the domain, $c'(x) \approx 0$ and there is little growth. Near $x = 0.5$, there is significant amplification, as shown in the plots of $c(x)$ and $s(x)$ in Fig. 1. The maximum growth in a one time unit is $\sigma_1(\mathbf{A}) = 1.17$. Calculation of the diagonal elements of \mathbf{B} and \mathbf{B}_T gives $\|\mathcal{L}_\mathbf{A}\|_\infty = \lambda_1(\mathbf{B}) = 347.1$ and $\|\mathcal{L}_\mathbf{A}\|_1 = \lambda_1(\mathbf{B}_T) = 334.0$. The spatial dependence of the diagonal elements of \mathbf{B} and \mathbf{B}_T is plotted in Fig. 2. From (41), the operator norm $\|\mathcal{L}_\mathbf{A}\|_p$ is large because the distance $r(\mathbf{A})$ from \mathbf{A} to the closest unstable matrix is small. If the matrix \mathbf{A} were normal, then $r(\mathbf{A})$ would be $1 - |\lambda_1(\mathbf{A})| = 1$. For the nonnormal dynamics here, the eigenvalues of \mathbf{A} are sensitive to perturbations and $r(\mathbf{A}) = 0.012$. The norm of $\mathcal{L}_\mathbf{A}^{-1}$ is bounded by $1 \leq \|\mathcal{L}_\mathbf{A}^{-1}\|_p \leq 1 + \sigma_1^2(\mathbf{A}) = 2.36$ from (40). Since the product $\|\mathcal{L}_\mathbf{A}\|_\infty \|\mathcal{L}_\mathbf{A}^{-1}\|_\infty$ is large, (39) tells us that the steady-state covariance matrix \mathbf{P} may have a good low-rank approximation.

To illustrate the bound matrix estimates derived from (47), we take the forcing covariance to be $\mathbf{P}_0 = 0.5\mathbf{I} + 0.5\mathbf{G}$, where \mathbf{G} is a Gaussian covariance model with correlation length 0.25 and normalized so that $\text{tr } \mathbf{G} = \text{tr } \mathbf{I}$; the diagonal elements of \mathbf{P}_0 are unity. A sense of the temporal behaviour of the system is seen by looking at the spatial mean of a realization of the forcing ξ_k and of the response ϵ_k in Fig. 3. The dynamics, with no modal growth, amplifies the forcing and increases the time coherence.

Figure 4a shows the eigenvalues $\lambda_i(\mathbf{P})$ along with the upper and lower bounds obtained from the bound matrix \mathbf{B} and (44). Much of the variance of \mathbf{P} is contained in just the first few modes, as suggested by the ill-conditioning of $\mathcal{L}_\mathbf{A}$. The spread in the bounds is due to bounding \mathbf{P}_0 , whose spectrum is not flat, with multiples of the identity matrix. Figure 4b shows the diagonal of \mathbf{P} and its bounds obtained from the bound matrix and (45).

The fraction of total variance explained by the best projection $\mathbf{P}^{(r)}$ defined in (2) is shown as a function of r in Fig. 5; the first 10 eigenmodes of \mathbf{P} explain about 70% of the total variance. In addition, the fraction of total variance explained by other rank- r approximations of \mathbf{P} is shown, in particular,

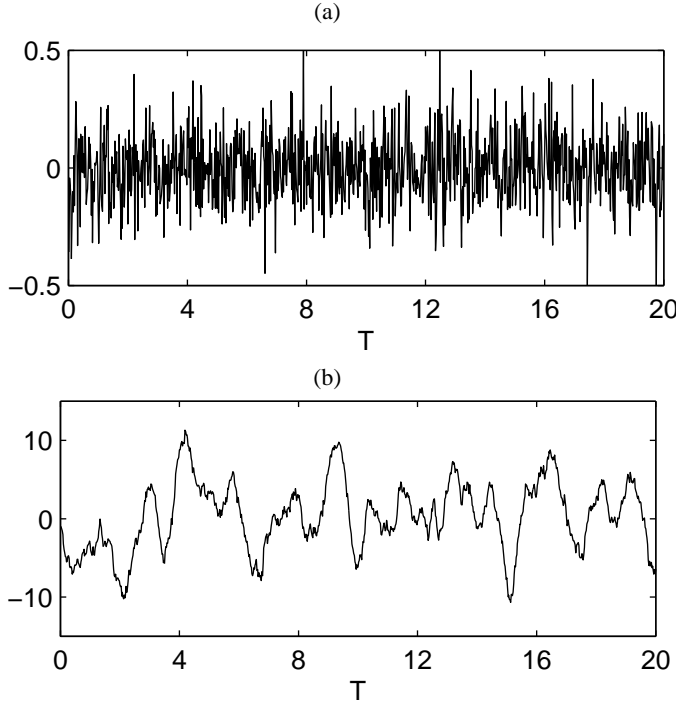


Fig. 3. Spatial mean of a realization of (a) the forcing ξ_k and (b) the response ϵ_k as a function of the advection time $T \equiv k\tau a$.

the projections onto the eigenvectors of \mathbf{B} , the left singular vectors of \mathbf{A} and the right singular vectors of \mathbf{A} . About 15 eigenvectors of \mathbf{B} are needed to explain 70% of the total variance. The left singular vectors of \mathbf{A} do not do as well, but are slightly better than the right singular vectors of \mathbf{A} .

5 Conclusions

Ensemble and reduced-state approaches to prediction and data assimilation have shown low-rank covariance representations to be practical covariance models. How appropriate such approximations are in a given problem depends on the spectrum of the full covariance, which is generally not available. Idealized and theoretical results must, therefore, be relied upon for guidance. We have obtained theoretical results in the case the covariance \mathbf{P} is a linear transformation \mathcal{L} of a forcing covariance \mathbf{P}_0 .

The singular value decomposition of the transformation \mathcal{L} provides information about the appropriateness of low-rank approximations of \mathbf{P} in some special cases, such as time-invariant systems with normal dynamics. More generally, we have seen that such information can be obtained from operator norms of the transformation \mathcal{L} . Since \mathcal{L} is a *positive map*, mapping covariance matrices to covariance matrices, there are simple expressions for its norms. Ill-conditioning is a necessary condition for the covariance matrix to permit a low-rank approximation. We have shown that the covariance \mathbf{P} can be ill-conditioned only when either the forcing covariance \mathbf{P}_0 or the transformation \mathcal{L} is also ill-conditioned.

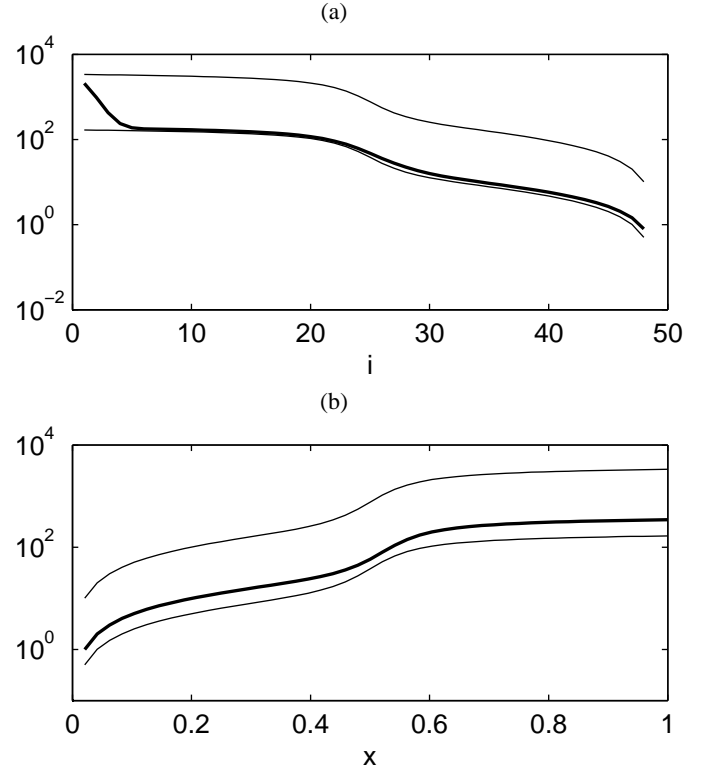


Fig. 4. (a) Eigenvalues $\lambda_i(\mathbf{P})$ (thick line) and their upper and lower bounds (thin lines) obtained using (44) with $\mathbf{P}_0^- = \lambda_n(\mathbf{P}_0)\mathbf{I}$ and $\mathbf{P}_0^+ = \lambda_1(\mathbf{P}_0)\mathbf{I}$. (b) Diagonal of \mathbf{P} (thick line) and its bounds (thin lines) obtained using (45) with $\mathbf{P}_0^- = \lambda_n(\mathbf{P}_0)\mathbf{I}$ and $\mathbf{P}_0^+ = \lambda_1(\mathbf{P}_0)\mathbf{I}$.

Similarly, we have shown that the fraction of the variance explained by the first eigenmode of the covariance can be large only when the same is true of the forcing covariance, or when \mathcal{L} is ill-conditioned.

In the case of the discrete algebraic Lyapunov equation describing the steady-state covariance matrix \mathbf{P} of a stable, time-invariant, stochastically forced dynamical system, conditioning of $\mathcal{L}_\mathbf{A}$ is related to the stability of the dynamics matrix \mathbf{A} . Nearly unstable dynamics leads to $\mathcal{L}_\mathbf{A}$ being ill-conditioned. These results were illustrated in an example using nonnormal dynamics from a generalized one-dimensional advection equation. The dynamics matrix \mathbf{A} is nearly unstable and $\mathcal{L}_\mathbf{A}$ is ill-conditioned. Analytic and numerical calculations of the singular values of \mathbf{A} , the radius of stability of \mathbf{A} and operator norms of $\mathcal{L}_\mathbf{A}$ demonstrated the mechanisms that lead to the ill-conditioning of $\mathcal{L}_\mathbf{A}$ and to the existence of good low-rank approximations of \mathbf{P} .

Appendix

Let \mathbf{A}_τ denote the operator that advances the solution of (49) from any time t to some later time $t + \tau$, $\epsilon(x, t + \tau) = \mathbf{A}_\tau \epsilon(x, t)$. The solution operator \mathbf{A}_τ can be determined explicitly by making a linear change in the variable such that in the new variable, (49) is just the constant-coefficient advection equation. The new dependent variable is $\nu(x, t) \equiv$

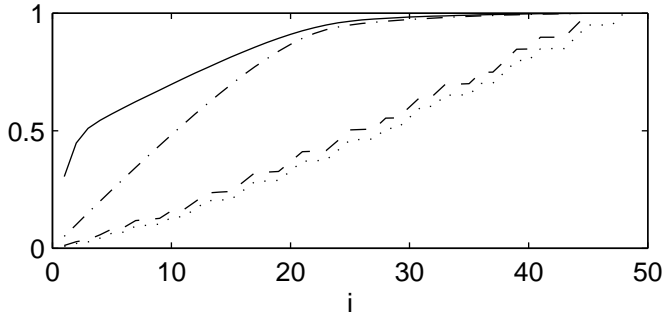


Fig. 5. Fraction of the total variance explained by the eigenvectors of \mathbf{P} (solid line), eigenvectors of \mathbf{B} (dash-dotted line), left singular vectors of \mathbf{A} (dashed line) and right singular vectors of \mathbf{A} (dotted line).

$L(x)\epsilon(x, t)$, where $L(x) \equiv \exp(c(x)/a)$, and satisfies

$$\nu_t + a\nu_x = 0 \quad (\text{A1})$$

with the initial and boundary conditions

$$\nu(x, t=0) = L(x)\epsilon_0(x), \quad \nu(x=0, t) = 0. \quad (\text{A2})$$

The solution operator $\tilde{\mathbf{A}}_\tau$ of this problem is given by

$$\tilde{\mathbf{A}}_\tau f(x) = \begin{cases} 0 & 0 \leq x \leq a\tau \\ f(x - a\tau) & a\tau \leq x \leq 1 \end{cases} \quad (\text{A3})$$

for functions $f(x)$ satisfying the boundary condition $f(0) = 0$. It follows from $\mathbf{A}_\tau = (L(x))^{-1} \tilde{\mathbf{A}}_\tau L(x)$ that \mathbf{A}_τ is given explicitly by

$$\mathbf{A}_\tau f(x) = \begin{cases} 0 & 0 \leq x \leq a\tau \\ s(x)f(x - a\tau) & a\tau \leq x \leq 1 \end{cases}, \quad (\text{A4})$$

where

$$s(x) \equiv (L(x))^{-1} L(x - a\tau) = \exp\left[\frac{c(x - a\tau) - c(x)}{a}\right]. \quad (\text{A5})$$

The solution operators $\tilde{\mathbf{A}}_\tau$ and \mathbf{A}_τ can be discretized exactly on a regular grid $\{x_1, x_2, \dots, x_n\}$, $x_i = i/n$, for integer Courant number $C \equiv a\tau n$. The $n \times n$ shift matrix $\tilde{\mathbf{A}}$ is defined as

$$\tilde{\mathbf{A}}_{ij} \equiv \begin{cases} 0 & 1 \leq i \leq C \\ \delta_{i-C, j} & C+1 \leq i \leq n \end{cases}, \quad (\text{A6})$$

where δ_{ij} is the Kronecker delta, and \mathbf{y} is a vector whose components are $y_i = f(x_i)$ for some function $f(x)$ defined on the grid. Then

$$(\tilde{\mathbf{A}}\mathbf{y})_i = \begin{cases} 0 & 1 \leq i \leq C \\ f(x_i - a\tau) & C+1 \leq i \leq n \end{cases} \quad (\text{A7})$$

and hence, $\tilde{\mathbf{A}}$ discretizes exactly the solution operator $\tilde{\mathbf{A}}_\tau$ defined in (A3). Similarly, the $n \times n$ matrix \mathbf{A} defined by

$$\mathbf{A}_{ij} \equiv \begin{cases} 0 & 1 \leq i \leq C \\ s(x_i)\delta_{i-C, j} & C+1 \leq i \leq n \end{cases} \quad (\text{A8})$$

discretizes the operator \mathbf{A}_τ in (A4) exactly. The matrices \mathbf{A} and $\tilde{\mathbf{A}}$ are related by $\mathbf{A} = \mathbf{L}^{-1} \tilde{\mathbf{A}} \mathbf{L}$, where \mathbf{L} is the diagonal matrix whose diagonal elements are $L_{ii} = L(x_i)$.

We want to calculate the singular values, singular vectors and bound matrix for the dynamics matrix \mathbf{A} . It follows from (A6) that $\tilde{\mathbf{A}} \mathbf{D} \tilde{\mathbf{A}}^T$ is diagonal for any diagonal matrix \mathbf{D} , with diagonal elements

$$(\tilde{\mathbf{A}} \mathbf{D} \tilde{\mathbf{A}}^T)_{ii} = \begin{cases} 0 & 1 \leq i \leq C \\ \mathbf{D}_{i-C, i-C} & C+1 \leq i \leq n. \end{cases} \quad (\text{A9})$$

Therefore, the matrix $\mathbf{A} \mathbf{A}^T = \mathbf{L}^{-1} \tilde{\mathbf{A}} \mathbf{L}^2 \tilde{\mathbf{A}}^T \mathbf{L}^{-1}$ is diagonal, with diagonal elements

$$(\mathbf{A} \mathbf{A}^T)_{ii} = \begin{cases} 0 & 1 \leq i \leq C \\ s^2(x_i) & C+1 \leq i \leq n. \end{cases} \quad (\text{A10})$$

Similarly, the matrix $\mathbf{A}^T \mathbf{A}$ is diagonal, with diagonal elements

$$(\mathbf{A}^T \mathbf{A})_{ii} = \begin{cases} s^2(x_i + a\tau) & 1 \leq i \leq n - C \\ 0 & n - C + 1 \leq i \leq n. \end{cases} \quad (\text{A11})$$

Thus, the singular values of \mathbf{A} are zero and the values of $s(x_i)$ for $i = C+1, \dots, n$. The left singular vector associated with the singular value $s(x_i)$ is the i -th column of the identity matrix, according to (A10), and the corresponding right singular vector is the $(i - C)$ -th column of the identity matrix, according to (A11).

The bound matrix \mathbf{B} satisfies $\mathbf{B} = \mathbf{A} \mathbf{B} \mathbf{A}^T + \mathbf{I}$, or equivalently

$$\tilde{\mathbf{B}} = \tilde{\mathbf{A}} \tilde{\mathbf{B}} \tilde{\mathbf{A}}^T + \mathbf{L}^2, \quad (\text{A12})$$

where $\tilde{\mathbf{B}} = \mathbf{L} \mathbf{B} \mathbf{L}$. It follows from (A9) that $\tilde{\mathbf{B}}$ is diagonal, with diagonal elements given by

$$\tilde{\mathbf{B}}_{ii} = \begin{cases} \mathbf{L}_{ii}^2 & 1 \leq i \leq C \\ \tilde{\mathbf{B}}_{i-C, i-C} + \mathbf{L}_{ii}^2 & C+1 \leq i \leq n. \end{cases} \quad (\text{A13})$$

Since $\tilde{\mathbf{B}}$ is diagonal, $\mathbf{B} = \mathbf{L}^{-1} \tilde{\mathbf{B}} \mathbf{L}^{-1}$ is also diagonal. For Courant number $C = 1$, equation (A13) is a first order difference equation whose solution is

$$\tilde{\mathbf{B}}_{ii} = \sum_{k=1}^i \mathbf{L}_{kk}^2 \quad (\text{A14})$$

for $i = 1, 2, \dots, n$, so that

$$\mathbf{B}_{ii} = \mathbf{L}_{ii}^{-2} \sum_{k=1}^i \mathbf{L}_{kk}^2 = \sum_{k=1}^i \exp\left[-\frac{2}{a}(c(x_i) - c(x_k))\right] \quad (\text{A15})$$

for $i = 1, 2, \dots, n$. Further, if $C = 1$, then (A13) gives

$$\mathbf{B}_{ii} = \frac{\mathbf{L}_{i-1,i-1}^2}{\mathbf{I}_{ii}^2} \mathbf{B}_{i-1,i-1} + 1 > \frac{\mathbf{L}_{i-1,i-1}^2}{\mathbf{I}_{ii}^2} \mathbf{B}_{i-1,i-1} \quad (\text{A16})$$

for $i = 2, 3, \dots, n$. If $c(x)$ is non-increasing, as assumed in the text, then $L(x)$ is also non-increasing and therefore, (A16) implies that $\mathbf{B}_{ii} > \mathbf{B}_{i-1,i-1}$; in particular, the maximum diagonal element of \mathbf{B} is at the right boundary $i = n$. Similarly, it can be shown that for $C > 1$, the maximum still occurs within $a\tau$ of the right boundary. The solution \mathbf{B}_T of $\mathbf{B}_T = \mathbf{A}^T \mathbf{B}_T \mathbf{A} + \mathbf{I}$ is also diagonal and for $C = 1$, the solution has diagonal elements given by

$$(\mathbf{B}_T)_{ii} = \mathbf{L}_{ii}^2 \sum_{k=i}^n \mathbf{L}_{kk}^{-2} = \sum_{k=i}^n \exp \left[\frac{2}{a} (c(x_i) - c(x_k)) \right] \quad (\text{A17})$$

for $i = 1, 2, \dots, n$; if $c(x)$ is non-increasing, then $(\mathbf{B}_T)_{ii} > (\mathbf{B}_T)_{i+1,i+1}$ and the maximum diagonal element of \mathbf{B}_T occurs at $i = 1$. Similarly, it can be shown that for $C > 1$, the maximum still occurs within $a\tau$ of the left boundary.

Acknowledgements. This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) Grants 91.0029/95-4, 381737/97-7 and 30.0204/83-3, Financiadora de Estudos e Projetos (FINEP) Grant 77.97.0315.00, and the NASA EOS Interdisciplinary Project on Data Assimilation. The authors give special thanks to D. Marchesin for his support and to P. Chang for his useful comments. The reviews of J. Hansen of MIT and an anonymous reviewer helped to improve the presentation of this work.

References

- Barkmeijer, J., Gijzen, M. V., and Bouttier, F.: Singular vectors and estimates of the analysis-error covariance matrix, *Q. J. R. Meteorol. Soc.*, 124, 1695–1713, 1998.
- Bhatia, R.: A Note on the Lyapunov Equation, *Linear Algebra Appl.*, 259, 71–76, 1997.
- Byers, R. and Nash, S.: On the singular “vectors” of the Lyapunov operator, *SIAM J. Alg. Disc. Math.*, 8, 59–66, 1987.
- Cane, M. A., Kaplan, A., Miller, R., Tang, B., Hackert, E., and Busalacchi, A. J.: Mapping tropical Pacific sea level: Data assimilation via a reduced state space Kalman filter, *J. Geophys. Res.*, 101, 22,599–22,617, 1996.
- Chang, P., Link, J., and Saravanan, R.: A Hybrid Coupled Model Study of Tropical Atlantic Variability, *J. Clim.*, 14, 3610–390, 2001.
- Changnon, D., Creech, T., and Marsili, N.: Interactions with a weather-sensitive decision maker: A case study incorporating ENSO information into a strategy for purchasing natural gas, *Bull. Amer. Meteor. Soc.*, 80, 1117–1126, 1999.
- Cohn, S. E.: An introduction to estimation theory, *J. Meteor. Soc. Japan*, 75, 257–288, 1997.
- Cohn, S. E. and Todling, R.: Approximate data assimilation schemes for stable and unstable dynamics, *J. Meteor. Soc. Japan*, 74, 63–75, 1996.
- Dee, D. P. and da Silva, A. M.: Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology, *Mon. Wea. Rev.*, 1998.
- Ehrendorfer, M. and Tribbia, J.: Optimal Prediction of Forecast Error Covariances through Singular Vectors, *J. Atmos. Sci.*, 54, 286–313, 1997.
- Farrell, B. F. and Ioannou, P. J.: Generalized stability theory. Part I: Autonomous operators, *J. Atmos. Sci.*, 53, 2025–2040, 1996.
- Ghivimi, A. R. and Laub, A. J.: Computation of approximate null vectors of Sylvester and Lyapunov operators, *IEEE Trans. Automat. Contr.*, 40, 387–391, 1995.
- Golub, G. H. and Van Loan, C. F.: *Matrix Computations*, The Johns Hopkins University Press, Baltimore, Third edn., 694 pp., 1996.
- Hamill, T. M. and Snyder, C.: A hybrid ensemble Kalman filter/3D-variational analysis scheme, *Mon. Wea. Rev.*, 128, 2905–2919, 2000.
- Horn, R. A. and Johnson, C. R.: *Matrix Analysis*, Cambridge University Press, New York, 561 pp., 1985.
- Houtekamer, P. L. and Mitchell, H. L.: Data Assimilation Using an Ensemble Kalman Filter Technique, *Mon. Wea. Rev.*, 126, 796–811, 1998.
- Kleeman, R. and Moore, A. M.: A theory for the limitation of ENSO predictability due to stochastic atmospheric transients, *J. Atmos. Sci.*, 54, 753–767, 1997.
- Lancaster, P.: Explicit solutions of linear matrix equations, *SIAM Review*, 12, 544–566, 1970.
- Molteni, F., Buizza, R., Palmer, T. N., and Petroliagis, T.: The ECMWF ensemble prediction system: Methodology and validation, *Q. J. R. Meteorol. Soc.*, 122, 73–119, 1996.
- Mori, T.: On the relationship between the spectral radius and the stability radius of discrete systems, *IEEE Trans. Automat. Contr.*, 35, 835, 1990.
- Rabier, F., McNally, A., Andersson, E., Courtier, P., Unden, P., Eyre, J., Hollingsworth, A., and Bouttier, F.: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: Structure functions, *Q. J. R. Meteorol. Soc.*, 124, 1809–1829, 1998.
- Riishøjgaard, L. P.: A direct way of specifying flow-dependent background error correlations for meteorological analysis systems, *Tellus*, 50A, 42–57, 1998.
- Tippett, M. K. and Marchesin, D.: Bounds for solutions of the discrete algebraic Lyapunov equation, *IEEE Trans. Automat. Contr.*, 44, 214–218, 1999.
- Tippett, M. K., Cohn, S. E., Todling, R., and Marchesin, D.: Conditioning of the stable, discrete-time Lyapunov operator, *SIAM J. Matrix Anal. Appl.*, 22, 56–65, 2000a.
- Tippett, M. K., Cohn, S. E., Todling, R., and Marchesin, D.: Low-dimensional representation of error covariance, *Tellus*, 52, 533–553, 2000b.
- Trefethen, L. N.: Pseudospectra of linear operators, *SIAM Review*, 39, 383–406, 1997.
- Whitaker, J. S. and Sardeshmukh, P. D.: A linear theory of extratropical synoptic eddy statistics, *J. Atmos. Sci.*, 55, 237–258, 1998.